Supplemental File: Televised Medical Talk Shows

Index

Additional Methods Notes

We attempted to contact The Dr. Oz Show to determine if transcripts and/or DVDs were available, but did not receive a response. Of the 40 randomized episodes, one episode was interrupted by breaking news, so a second one was randomly selected. For The Doctors, three episodes were interrupted by other programming, and two were repeats, thus five additional episodes were randomly selected.

Data extraction

Data extracted from each episode included the following: Date of show airing, educational topics discussed, key points made on the show, specific recommendations, who the recommendations were for (e.g. everyone vs specific populations), who made the recommendation (e.g. host vs guest), the presence of written information on the screen to reinforce recommendations, the use of testimonials, references to scientific studies and citations if mentioned, promotional placements, guests on the show including any information regarding their area of specialty, perceived negative comments about healthcare professionals, discussion of provocative topics, and the use of medical jargon.

In the second round of viewing, additional data specific to the recommendations was extracted including: was a benefit mentioned in regards to the recommendation (or harm)? Was the magnitude of the benefit mentioned? Were potential harms mentioned? Was cost of the recommendation discussed? For each item we recorded the actual benefit, magnitude, harm and or cost reported. We also recorded acknowledgment of potential conflict of interest.

Agreement between reviewers was calculated for all extracted data.

Categorization of Topics

All educational topics were categorized initially by the first two reviewers into one of the following categories: weight loss, alternative therapies, vitamins, screening/prevention, beauty and appearance, exercise, physical non-medical, provocative topics, time filler/random topic, gossip, illness/disease, food, health technology, medical innovations and news. Further review suggested that fewer categorizations would be more informative in giving an overall

impression of what topics are being discussed. The topics were recoded (CK, JM) into the following: weight loss, counseling (e.g. romance recommendations), general medical advice (e.g. allergies, health news, embarrassing body questions), non-medical (e.g. embarrassing pet behaviour), dietary advice not related to weight loss (e.g. diets to improve fatigue, health), alternative therapy (e.g. home remedies), cosmetic, exercise and other.

Categorization of Recommendations

Recommendations were initially coded according to the following categories: screening/prevention, food, tests, alternative supplements, physical therapy, drugs, vitamins, consult, treatment, behavior, remedy and exercise. As described in the protocol change, these were recategorized in an attempt to more clearly reflect the content of the recommendations. The modified categories included: alternative oral therapy (e.g. use 5-HTP in practice today to suppress appetite and increase mood), alternative non-oral therapy (e.g. can use mouth wash to combat dandruff - rub in hair with conditioner), consult a health care provider, cosmetic, counseling (e.g. wear red to give you energy and get you going in the morning), dietary advice (e.g. go get monk fruit sweetener/lo han sweetener and use it instead of sugar), medications (e.g. every women should take this supplement to reduce risk of heart disease: low dose ASA [2 x 81mg/ day]), drugs of abuse, dental care, diagnostic recommendations (e.g. bright red tongue may indicate anemia due to B12 deficiency), exercise, food preparation, herbs, homeopathy, lifestyle, motherhood statements (e.g. never leave home without clean underwear), infection prevention, non drug medical advice (e.g. use a wired ear piece for cell phone (keep phone away from brain and safer while driving)), physiotherapy, screening, sex related topics, tests, vitamins, vaccines and other.
Food preparation was reported as a subset of Dietary advice. Herbs and Homeopathy were reported as either oral or non-oral alternative therapy.

Selection of Recommendations for Evidence Analysis

As discussed in the main paper, more definitive recommendations were categorized as "stronger". Stronger recommendations were randomly selected for both The Doctors and The Dr. Oz Show with a target of 80 recommendations for each show. For The Doctors, we initially determined that 3 of the recommendations were not amenable to searching, due to the fact they were quite vague or did not have a clear outcome associated with them, (e.g. Don't

give up the foods you love, just prepare them in the healthiest way possible). In these cases 3 additional recommendations were randomized. For The Dr. Oz Show 16 recommendations were removed for similar reasons (e.g. Never leave home without clean underwear) or because of duplication of recommendations. Additional randomized recommendations were used to ensure we started with 80 recommendations in both groups.

Evidence analysis data template.

To assess whether recommendations were supported by evidence, we developed a searchable question for each randomly selected recommendation (RT, MA, CK). We recorded what would be the ideal form of evidence to support each recommendation (e.g. cohort or RCT). We also recorded whether the recommendation was based on clinical or surrogate outcomes (e.g for the question "Does fiber improve glycemic control?" glycemic control was considered a surrogate outcome). The best level of evidence found was recorded and the result of this evidence was reported as either "evidence of effect", "evidence of no effect", "evidence of harm", "no evidence found". For one recommendation from The Doctors, the group concluded that no evidence was needed as the recommendation was felt to be more of a definition of a psychiatric disorder (If you sit down and binge on 10-15000 calories, you need to get help to become healthier because that is a disorder). This recommendation was removed from the final calculations. One recommendation from The Doctors was felt to be two recommendations, so the final number of recommendations reviewed was consistent at 80. Following a review of the evidence, the group determined whether the recommendation was supported by evidence with either "yes" or "no" being reported. A case report was considered the lowest level of evidence acceptable to support a recommendation.

When reviewing evidence we frequently used a relatively broad definition of support in attempt to be as fair as possible. In many instances, surrogate markers were the only thing discussed. Without evidence proving that surrogate markers were incorrect, we were willing to accept evidence of change in surrogate markers. For instance, one recommendation on the Dr. Oz show was that you "want your CRP number to be less than 1.0mg/L." This would imply that you have to change your CRP. There is no evidence to suggest that changing your CRP makes a difference clinically and we don't have evidence on what numbers are best. However, there is evidence to suggest that elevated CRP can be associated with vascular

outcomes, so we rated this recommendation as having consistent and believable evidence to support it.

Another example, from one recommendation to “Eat green bananas”, came with the reported benefit that green bananas contain less sugar than ripe bananas and may ultimately help in the prevention of diabetes. One study showed a higher glucose load with ripe bananas but no difference in insulin levels1 whereas another showed no difference in serum glycemic or insulin levels.2 This was recorded as evidence present supporting the recommendation.

Following this we assessed the consistency and believability of the evidence as described in the main paper. Agreement between the four reviewers was recorded for each recommendation, and comments to expand on our decision were recorded for each recommendation.

References

1) Hermansen K, Rasmussen O, Gregersen S, Larsen S. Influence of ripeness of banana on the blood glucose and insulin response in type 2 diabetic subjects. Diabet Med. 1992;9:739-43.
2) Ercan N, Nuttall FQ, Gannon MC, Lane JT, Burmeister LA, Westphal SA. Plasma glucose and insulin responses to bananas of varying ripeness in persons with noninsulin-dependent diabetes mellitus. J Am Coll Nutr. 1993;12:703-9.

Protocol Adaptation

At the onset of this study we were largely unfamiliar with broadcast health information, particularly the non-specific nature of statements and recommendations given on syndicated medical/health television talk shows. This led to three small protocol changes/additions and one larger change.

1) Our small sample review of two episodes of each show included a season of The Dr. Oz Show that included a written banner on the screen when more definitive recommendations were given. Unfortunately, the more recent season that we reviewed no longer included this banner, making stronger recommendations more difficult to identify. In the end, we decided to continue to attempt to identify recommendations that were stronger and use those for determination of evidence to support recommendations. However, due to the subjective nature of defining the strength of the recommendations, we decided not to analyze these two groups of recommendations in any further detail.

2) After our first data extraction was complete, it became clear that we were not adequately capturing the level of detail and information associated with each recommendation. We therefore added a second review to determine additional information was given regarding the benefits, risks and potential financial costs associated with each recommendation.

3) The first data extractors originally classified and coded all topics and recommendations. After review, we found they were not adequately descriptive or too broad. For example, one recommendation category was "remedy" meaning a product that could be used for something else (e.g. use lemons to provide a chemical peel for age spots). Another was "behavior" which could explain any activity from ordering a kids meal for weight loss to the right place to sit in the ER to make sure you're not forgotten. Therefore, two of primary investigators (CK, JM) went through and reclassified all topics and recommendations, with any uncertainties resolved by a third investigator (GMA).

4) Our original design for evidence analysis involved two reviewers working independently searching for and assessing if evidence supported a recommendation. Once reviewers were complete, the lead investigator (CK) compiled the data and determined agreement between the reviewers. In cases where reviewers did not agree,

the lead investigator (CK) reviewed all citations recorded by both reviewers and came to a final decision. Unfortunately, this presented multiple challenges. We realized that despite an attempt to focus on stronger recommendations, many remained quite broad and unfocused. Creating appropriate, searchable questions was quite difficult because we wanted the search question to reflect the recommendation, meaning we could not modify the wording much at all. Thus it was often difficult to define the population, intervention, or outcome of the question without making assumptions and potentially modifying the intent of the recommendation. In turn, this made identifying or determining whether the evidence supported recommendation even more difficult.

In the end, reviewer's interpretation of the evidence varied - initial data from The Dr. Oz Show found that reviewers agreed on the presence of evidence 63.8% of the time. This was based on a number of factors. First, some of the recommendations were extremely vague and the benefit or harm associated with the recommendation was unclear or not stated at all. Second, if the benefit was mentioned it could also be ambiguous (e.g. increases brain power). Third, some of the recommendations had more than one component (e.g. don't drink coffee because it will cause insomnia and increase your risk of an accident). Fourth, reviewers also disagreed on what was acceptable evidence (e.g. surrogate data for patient oriented outcomes, the presence of indirect evidence, the quality of the study). This was compounded in part by our directives given around searching for patient-oriented outcomes when some of the recommendations were entirely on surrogate markers (e.g. to reduce cholesterol). In some instances, there was evidence but the reviewer did not feel it was believable and thus concluded the evidence did not support the recommendation.

Finally, in order to more reliably determine whether the recommendations were supported by reasonable quality scientific evidence, it was determined that all recommendations should be consistently reviewed by a group of four experienced evidence reviewers (CK, MRK, JM, GMA). Infrequently one of the group (JM) was unavailable but the remainder of the group proceeded. Each recommendation and the evidence that had previously been located was reviewed in sequence by the four experienced evidence reviewers as a group. The group first reviewed the recommendation. If the recommendation included multiple ideas (e.g. Eat the following detox foods to look good and lose weight: 1. green leafy and cruciferous vegetables; 2. cold water fish; 3. free range pastured chicken…) only the first one

would be searched. The reviewers recorded what would be considered ideal evidence. Ultimately the presence of a case report or higher level of evidence was recorded as evidence being present. Expert opinion was not considered evidence. The group reviewed all available evidence. In some cases where the answer remained unclear, they would spend time completing another search of the literature until the group was satisfied that a reasonable answer had been found or could not be found. The group would discuss each recommendation and come to a consensus regarding whether available evidence supported the recommendation. In some cases, evidence was available to support the recommendation but the group did not feel it was consistent or believable. These outcomes were included in the spreadsheet in an attempt to capture the nuances of the evidence. Consistency was defined as both internal (within the study) and external (between studies). Believability was based on the quality, quantity and type of evidence available evidence. Based on early observation of a difference between the two shows, the group also recorded how frequently "consult a healthcare provider" was mentioned within the 160 recommendations that had been randomized.